

HT1220 Report

RECORD INFORMATION

Gene ID:	1220
Sequence ID:	1220
Protein ID:	1220
Sequence name:	thrombospondin 1, alt. transcript 1
Genome:	nucleus
Taxon:	Homo sapiens
Locus:	1220
Common Name:	thrombospondin 1
Role ID:	40

Coding sequence length:	3513 nt
Transcript sequence length:	5722 nt
Expression data:	<u>THC201673</u>

ACCESSION DATA

HT1220 is derived from accessions(s):

SP:P07996 (THROMBOSPONDIN 1 PRECURSOR.)
GB:X04665 (Human mRNA for thrombospondin)
GB:X14787 (Human mRNA for thrombospondin)
GB:U12471 (thrombospondin-p50 {Homo sapiens})
GB:M99425 (Human thrombospondin mRNA, 3' end.)
PIR:G01478 (thrombospondin-p50 - human (fragment))
GB:U12471 (Human thrombospondin-1 gene, partial cds.)
GB:J04835 (Human thrombospondin gene, exons 1, 2 and 3.)
GB:M25631 (Homo sapiens (clone lambda-TS-33) thrombospondin (THBS) mRNA, 5' end.)

ALTERNATIVE SPLICE INFORMATION

Alternative splice forms for this gene:

HT3987 thrombospondin 1, alt. transcript 2

MAPPING DATA

GDB accession(s) for this gene:

GDB ID:	Symbol

Figure 1A

gdb:120438 THBS1

cDNA FEATURES

Feature	End 5	End 3
coding_seq	112	3624
3'UT	3625	5722
spjunc_h	1235	1236

SEQUENCE

nucleotide:

```

ggacgcacaggcattccccgcgcccctccagccctcgccgcccctcgccaccgctcccggc
cgccgcgctccgggtacacacaggatccctgctgggcaccaacagctccaccatggggctg
gcctggggactaggcgtcctgttccctgatgcattgtgtgtggcaccacacgcatccagag
tctggcgagagacaacagcgtgtttgacatccttgaactcaccggggcgcccgcaagggg
tctggggcgccgactggtgaaggggccccgaccctccagcccagctttccgcacagaggat
gccaacctgatccccctgtgcctgatgacaagtccaagacctggtggatgctgtgagg
gcagaaaagggtttcctccttctggcatccctgaggcagatgaagaagaccggggcagc
ctgctggccctggagcggaaagaccactctggccaggtcttcagcgtggtgtccaatggc
aaggcgggcaccctggacctcagcctgaccgtccaaggaaagcagcagctggtgtctgtg
gaagaagctctcctggcaaccggccagtggaagagcatcaccctgtttgtgcaggaagac
aggggccagctgtacatcgactgtgaaaagatggagaatgctgagttggacgtccccatc
caaagcgtcttcaccagagacctggccagcatcgccagactccgcacgcaaaaggggggc
gtcaatgacaatttccagggggtgctgcagaatgtgaggtttgtcttggaaaccacacca
gaagacatcctcaggaacaaaggctgctccagctctaccagtgctcctcacccttgac
aacaacgtggtgaatggttccagccctgccatccgcactaactacattggccacaagaca
aaggacttgcaagccatctgcggcatctcctgtgatgagctgtccagcatggtcctggaa
ctcaggggctgcgccaccattgtgaccacgctgcaggacagcatccgcaaagtgactgaa
gagaacaaagagttggccaatgagctgaggcgccctcccctatgctatcacaacggagtt
cagtacagaaaataacgaggaatggactgttgatagctgcactgagtgctactgtcagaac
tcagttaccatctgcaaaaagggtgtcctgcccccatcatgcccgtgctccaatgccacagtt
cctgatggagaatgctgtcctcgtgttggcccagcagctctgcggacgatggctggtct
ccatggtccgagtggaacctcctgttctacgagctgtggcaatggaattcagcagcgcggc
cgctcctgcgatagcctcaacaaccgatgtgagggtcctcgggtccagacacggacctgc
cacattcaggagtggtgacaaaagatttaaacaggatggtggctggagccactggtccccg
tggtcatcttgttctgtgacatgtggtgatggtgtgatcacaaggatccggctctgcaac
tctcccagccccagatgaatgggaaacctgtgaaggcgaagcgcgggagaccaaagcc
tgcaagaaagacgcctgcccccatcaatggaggctggggtccttggtcaccatgggacatc
tggtctgtcacctgtggaggaggggtacagaaacgtagtcgtctctgcaacaaccccgca
ccccagtttgagggcaaggactgcgttggtgatgtaacagaaaaccagatctgcaacaag
caggactgtccaattgatggatgcctgtccaatccctgctttgcccggctgaagtgtact
agctaccctgatggcagctggaaaatgtggtgcttgtccccctgggttacagtggaaatggc
atccagtgacagatgttgatgagtgcaagaagtgcctgatgcctgcttcaaccacaat
ggagagcacccggtgtgagaacacggacccccggctacaactgcctgcctgccccccacgc
ttcaccggctcacagcccttcggccagggtgtcgaaatgccaacggccaacaaacaggtg
tgcaagccccgtaacccctgcacggatgggacccacgactgcaacaagaacgcgaagtgc
aactacctggggccactatagcgaccccatgtaccgctgcgagtgcaagcctggctacgct
ggcaatggcatcatctgcggggaggacacagacctggatggctggcccaatgagaacctg
gtgtgctggccaatgcgacttaccactgcaaaaaggataattgccccaaccttcccaac
tcagggcaggaagactatgacaaggatggaattgggtgatgcctgtgatgatgacgatgac
aatgatataaattccagatgacagggacaactgtccattccattacaaccagctcagtat
gactatgacagagatgatgtgggagaccgctgtgacaactgtccctacaaccacaaccca
  
```

Figure 1B

gatcaggcagacacagacaacaatggggaaggagacgcctgtgctgcagacattgatgga
gacgggtatcctcaatgaacgggacaactgccagtagctctacaatgtggaccagagagac
actgatatggatgggggttgagatcagtgtagacaattggcccttggaacacaatccggat
cagctggactctgactcagaccgcattggagatacctgtgacaacaatcaggatattgat
gaagatggccaccagaacaatctggacaactgtccctatgtgccaatgccaaccaggct
gaccatgacaaaagatggcaaggagatgcctgtgaccacgatgatgacaacgatggcatt
cctgatgacaaggacaactgcagactcgtgccaatcccgaccagaaggactctgacggc
gatggtcgaggtgatgcctgcaaagatgattttgaccatgacagtggtgccagacatcgat
gacatctgtcctgagaatgttgacatcagtgagaccgatttccgccgattccagatgatt
cctctggaccccaaaggacatcccaaatgaccctaactgggttgtagccatcagggt
aaagaactcgtccagactgtcaactgtgatcctggactcgtgtagggttatgatgagttt
aatgctgtggacttcagtggcaccttcttcatcaacaccgaaaggacgatgactatgct
ggattttgtctttggctaccagtcacgacggcgttttatgttgtgatgtggaagcaagtc
accagtcctactgggacaccaacccacgagggctcagggatactcgggcctttctgtg
aaagtgttaaactccaccacagggcctggcgagcacctgcggaacgcctgtgggcacaca
ggaaaacacccctggccaggtgcgacccctgtggcatgaccctcgtcacataggctggaaa
gatttcaccgcctacagatggcgtctcagccacaggccaaagacgggtttcattagagt
gtgatgtatgaagggaagaaaatcatggctgactcaggaccatctatgataaaacctat
gctgggtgtagactagggttggtttgtcttctctcaagaaatgggtgttcttctctgacctg
aaatacgaatgtagagatccctaatacatcaaatgttgattgaaagactgatcataaacc
aatgctggtattgcaccttctggaactatgggcttgagaaaaacccaggatcacttctc
cttggcttcttcttcttctgtgcttgcatcagtgtagactcctagaacgtgcgacctgcc
tcaagaaaatgcagttttcaaaaacagactcatcagcattcagcctccaatgaataagac
atcttccaagcatataaaacaattgctttgggttcttctttgaaaaagcatctacttgcttc
agttgggaagggtgccattccactctgcctttgtcacagagcagggtgctatttgtgaggc
catctctgagcagtgtagctcaaaagcattttcaggcatgtcagagaaggaggactcact
agaattagcaaaacaaaaccacctgacatcctccttcaggaacacggggagcagaggcca
aagcactaagggggagggcgcataccgcagacgattgtatgaagaaaatatggaggaactg
ttacatgttccggtactaagtcattttcaggggattgaaagactattgctggatttcatga
tgctgactggcgttagctgattaacccatgtaaataggcacttaaatagaagcaggaaaag
ggagacaaaagactggcttctggacttccctcctgatccccacccttactcatcacttgc
agtggccagaattaggggaatcagaatcaaacacagtgtaaggcagtgctggctgccattgc
ctggctcacattgaaattgggtggcttcatttctagatgtagcttgtgcagatgtagcaggaa
aataggaaaacctaccatctcagtgagcaccagctgcctcccaaaggaggggcagccgtg
cttatattttttatgggttacaatggcacaaaattattatcaacctaaactaaaacattcctt
ttctcttttttccgtaattactaggtagttttcctaattctctcttttggaaagtatgattt
ttttaaagtctttacgatgtaaaatatttatttttacttattctggaagatctggctga
aggattattcatggaacaggaagaagcgtaaagactatccatgtcatctttgttgagagt
cttcgtgactgtaagattgtaaatacagattatttatttaactctgttctgcctggaaatt
taggcttcatacggaaaagtgttgagagcaagtagttgacatttatcagcaaatctcttg
caagaacagcacaaaggaaaatcagtcataaagctgctctgccccttgctcagagtgg
atgttatgggattccttttttctctgttttatcttttcaagtggaaattagttgggttatcc
atttgcaaatgttttaaatgcaaaagccatgaggtcttcaatactgttttacccca
tcccttgctcatatttccaggggagaaggaaagcatatacacttttttctttcatttttcc
aaaagagaaaaaaatgacaaaagggtgaaacttacatacaaatattacctcatttgttgtg
tgactgagtaaaagaatttttggatcaagcggaaaagagtttaagtgtctaaacaaactaaa
gctactgtagtacctaaaaagtcagtggtgtacatagcataaaaaactctgcagagaagta
ttcccaataaggaaaatagcattgaaatgttaaatacaatttctgaaagttatgtttttt
tctatcatctgggtataccattgcttttatttttataaattattttctcattgccattggaa
tagaataattcagatttgtgtagatatgctattttaaataatttatcaggaataactgcctgt
agagttagttttctattttttatataatgtttgcacactgaattgaagaattgttgggtt
tttctttttttgtttttttttttttttttttttttgttttgacctccattttta
ctatttgccaataaccttttctaggaatgtgctttttttgtacacatttttatccattt
tacattctaaagcagtgtaagttgtatattactgtttcttatgtacaaggaacaacaata
aatcatatggaaaatttatattt

protein:

MGLAWGLGVFLMHVCGTNRIPESGGDNSVFDIFELTGAARKGSGRRLVKGPDPSSPAFR

Figure 1C

IEDANLIPPVPDDKFQDLVDAVRAEKGFLLLASLRQMCKTRGTLALERKDHSGQVFSVV
SNGKAGTLDLSLTVQKGQHVVSVEEALLATGQWKSITLFFVQEDRAQLYIDCEKMENAEILD
VPIQSVFTRDLASIALRLIAKGGVNDNFQGV LQNVRFVFGTTPEDILRNKGCSSSTSLL
TLDNNVVGSSSPAIRTNYIGHKTKDLQAICGISDELSSMVLELRGLRTIVTTLQDSIRK
VTEENKELANELRRPPLCYHNGVQYRNNEEWTVDSCTECHCQNSVTICKKVSCPIMPASN
ATVPDGECCPRCWPSPDSADDGWSPWSEWTSCSTSCGNGIQQRGRSCDSLNNRCEGSSVQT
RTCHIQECDKRFKQDGGWSHWSPWSSCSVTCDGVITRIRLCNSPSPQMNGKPCGEARE
TKACKKDACPINGGWGPWSPWDICSVTCGGGVQKRSRLCNPAPQFGGKDCVGDVTENQI
CNKQDCPIDGCLSNPCFAGVKCTSYPDGSWKCGACPPGYSGNGIQCTDVDECKEVPDACF
NHNGEHRCENTDPGYNCLPCPPRFTGSQPFQGVGHATANKQVCKPRNPCTDGTDCNKN
AKCNYLGHYSDPMYRCECKPGYAGNGIICGEDTDLGWPNEENLVCVANATYHCKKONCPN
LPNSGQEDYDKDGIGDACDDDDNDKI PDDRNCPPHYNPAQYDYDRDDVGDRCDCNCPYN
HNPDAQADTDNNGEGDACAADIDGDGILNERDNCQYVYNVDQRDTMDGVDGQCDNCPLEH
NPDQLDSDSDRIGDTCNNQDIDEDGHQNNLDNCPYVNPANQADHDKDGKGDACDHDDDN
DGIPDDKONCRLVNPDPQKSDGDGRGDACKDDFDHDSVPDIDDICPENVDISETDFRRF
QMIPLDPKGTSQNDPNWVVRHQKELVQTVNCDPGLAVGYDEFNAVDFSGTFFINTERDD
DYAGFVFGYQSSSRFYVVMWKQVTQSYWDTNPTRAQGYSGLSVKVNVNSTTGPEHLRNAL
WHTGNTPGQVRTLWHDPRHIGWKDFTAYRWRLSHRPKTGFIRVVMYEGKKIMADSGPIYD
KTYAGGRLGLFVFSQEMVFFSDLKYECRDP

Figure 1D

HT2143 Report

RECORD INFORMATION

Gene ID: 2081
Sequence ID: 2143
Protein ID: 2125
Sequence name: thrombospondin 4
Genome: nucleus
Taxon: Homo sapiens
Locus: 2081
Common Name: thrombospondin 4
Role ID: 40

Coding sequence length: 2886 nt
Transcript sequence length: 3074 nt
Expression data: THC168897

ACCESSION DATA

HT2143 is derived from accessions(s):

SP:P35443 (THROMBOSPONDIN 4 PRECURSOR.)
GB:Z19585 (thrombospondin-4 {Homo sapiens})
GB:Z19585 (H.sapiens mRNA for thrombospondin-4)
PIR:A55710 (thrombospondin 4 precursor - human)

cDNA FEATURES

Feature	End 5	End 3
coding_seq	28	2913
3'UT	2914	3074

SEQUENCE

nucleotide:

gaattccggggagcaggaagagccaacatgctggccccgcggagccgcccgtcctcctg
ctgcacctgggtcctgcagcgggtggctagcggcaggcgcccaggccacccccagggtctt
gaccttctcccatcttccagtcagagggttaaaccaggcgctctgctgccagtctgaca
gaccccgccctgaatgatctctatgtgatttccaccttcaagctgcagactaaaagttca
gccaccatcttcgggtctttactcttcaactgacaacagtaaatattttgaatttactgtg
atgggacgcttaagcaaagccatcctccgttacctgaagaacgatgggaagggtgcatttg

Figure 2A

gtgggttttcaacaacctgcagctggcagacggaaggcgccacaggatcctcctgaggctg
agcaatattgcagcggggccggctccctagagctctacctggactgcatccagggtggat
tccgttcacaatctccccagggcctttgctggccctcccagaaacctgagaccattgaa
ttgaggactttccagaggaagccacaggacttcttggaagagctgaagctgggtgaga
ggctcactgttccagggtggccagcctgcaagactgcttctgcagcagagtgcaccactg
gctgccacaggcacaggggactttaaccggcagttcttgggtcaaataacacaaat
caactcctgggagaggtgaaggaccttctgagacagcagggttaaggaaacatcatttttg
cgaaacaccatagctgaatgccaggcttgcggtcctctcaagtttcagttctccgacccca
agcacgggtgggtcgccccggctccccctgcaccgccaacacgcccacctcgtcggtgtgac
tccaacccatgtttccgaggtgtccaatgtaccgacagtagagatggcttccagtggtggg
ccctgccccgagggctacacaggaaacgggatcacctgtattgatgtgatgagtgcaaa
taccatccctgctaccggggcgtgcactgcataaatttgcctcctggcttcagatgtgac
gcttgcctcagtggtgttcacagggcccatgggtgcagggtgttgggatcagtttggcgaag
tcaaaccaagcagggtctgcactgacattgatgagtgctgaaatggagcgtgcgttcccaac
tcgatctgcgttaatactttgggatcttaccgctgtgggccttgaagccgggggtatact
gggtgatcagataaggggatgcaaagtggaaagaaactgcagaaaccagagctgaaccct
tgcagtgatgaatgccagtgcatgaagagaggcagggggatgtgacatgtgtgtgtgga
gtcgggtgggctggagatggctatatctgtggaaaggatgtggacatcgacagttacccc
gacgaagaactgccatgctctgccaggaactgtaaaaaggacaactgcaaatatgtgcca
aattctggccaagaagatgcagacagagatggcattggcgacgcttgtgacgaggtgct
gacggagatgggatcctgaatgagcaggataaactgtgtcctgattcataatgtggaccaa
aggaacagcgataaagatatctttggggatgcctgtgataactgcctgagtgctttaa
aacgaccagaaagacaccgatggggatggaagaggagatgcctgtgatgatgacatggat
ggagatggaataaaaaacattctggacaactgccccaaatttcccaatcgtgaccaacgg
gacaaggatgggtgatgggtgtgggggatgcctgtgacagttgtcctgatgtcagcaaccct
aaccagttctgatgtggataatgatctggttggggactcctgtgacaccaatcaggacagt
gatggagatgggacaccagagacagacagacaactgccccaccgtcattaacagtgccacg
ctggacaccgataaaggatggaattgggtgacgagtgatgatgatgatgacaatgatggt
atcccagacctgggtgccccctggaccagacaactgcccggctgggtccccaaccagcccag
gaggatagcaacagcgacggagtgaggagacatctgtgagttgactttgaccaggaccag
gtcatcgatcggtacgctctgcccagagaacgcagaggtcacccctgaccgacttcagg
gcttaccagaccgtgggcctggatcctgaaggggatgcccagatcgatcccaactgggtg
gtcctgaaccagggcatggagattgtacagaccatgaacagtgatcctggcctggcagtg
gggtacacagcttttaattggagttgacttcgaaggaccttccatgtgaataccagaca
gatgatgactatgcaggctttatctttgggtaccaagatagctccagcttctacgtggct
atgtggaagcagacggagcagacatatgggaagccacccattccgagcagttgcagaa
cctggcattcagctcaaggctgtgaagtctaagacaggtccaggggagcatctccggaac
tccctgtggcacacgggggacaccagtgcaccaggtcagggtgctgtggaaggactccagg
aatgttgggctggaaggacaagggtgtcctaccgctggttctacagcacaggccccagggtg
ggctacatcagggtacgattttatgaaggctctgagttgggtggtgactctggcgtcacc
atagacaccacaatgcgtggaggccgacttggcgttttctgcttctcaagaaaacatc
atctgggtccaacctcaagtatcgctgcaatgacaccatccctgaggacttccaagagttt
caaaccagaaatttcgaccgcttcgataattaaaccaaggaagcaatctgtaactgcttt
tcggaacactaaaaccatataattttaacttcaattttctttagcttttaccacccaa
atatatcaaaacgttttatgtgaatgtggcaataaaggagaagagatcatttttaaaaaa
aaaaaaaaaaaaaa

protein:

MLAPRGA AVL L L L H L V L Q R W L A A G A Q A T P Q V F D L L P S S S Q R L N P G A L L P V L T D P A L N D L Y V
I S T F K L Q T K S S A T I F G L Y S S T D N S K Y F E F T V M G R L S K A I L R Y L K N D G K V H L V V F N N L Q L A
D G R R H R I L L R L S N L Q R G A G S L E L Y L D C I Q V D S V H N L P R A F A G P S Q K P E T I E L R T F Q R K P Q
D F L E E L K L V V R G S L F Q V A S L Q D C F L Q Q S E P L A A T G T G D F N R Q F L G Q M T Q L N Q L L G E V K D L
L R Q Q V K E T S F L R N T I A E C Q A C G P L K F Q S P T P S T V V A P A P P A P P T R P P R R C D S N P C F R G V Q
C T D S R D G F Q C G P C P E G Y T G N G I T C I D V D E C K Y H P C Y P G V H C I N L S P G F R C D A C P V G F T G P
M V Q G V G I S F A K S N K V C T D I D E C R N G A C V P N S I C V N T L G S Y R C G P C K P G Y T G D Q I R G C K V
E R N C R N P E L N P C S V N A Q C I E E R Q G D V T C V C G V G W A G D G Y I C G K D V D I D S Y P D E E L P C S A R
N C K K D N C K Y V P N S G Q E D A D R D G I G D A C D E D A D G D G I L N E Q D N C V L I H N V D Q R N S D K D I F G
D A C D N C L S V L N N D Q K D T D G D G R G D A C D D D M D G D G I K N I L D N C P K F P N R D Q R D K D G D G V G D

Figure 2B

ACDSCPDVSNPNQSDVDNDLVGDS CDTNQSDGDGHQDSTDNCPTVINS AQLD TDKDGIG
DECDDDDNDGIPDLVPPGPDNCRLVPNPAQEDSNSDGVGDICESDFDQDQVIDRIDVCP
ENAEVTLTDFRAYQTVGLDPEGDAQIDPNWVVLNQGMEIVQTMNSDPGLAVGYTAFNGVD
FEGTFHVNTQTDDDYAGFIFGYQDSSSFYVVMWKQTEQTYWQATPFRAVAEPGIQLKAVK
SKTGPGEHLRNSLWHTGDTSDQVRLWLDKSRNVGWKDKVSYRWFLQHRPQVGYIRVRFYE
GSELVADSGVTIDTTMRGGRLGVFCFSQENIIWSNLKYRCNDTIPEDFQEFQTQNFDRFD
N

Figure 2C

Poly ID	Sequence ID	Pos Mon	Gene Description	Flanking Seq	Mutation T/Ref NT	AR NT	Ref AA	Alt AA
G334i4	1T11T1220_mRNA	2210	THBS1, Interleukin-1	TTGATGGCTGGCCCAJA/GTGAAGAACCTGGTGTG	Missense A	G	N	S
G335i2	1T11T12143_mRNA	1106	THBS4, Interleukin-4	GAGTGTGGAATGGAG/C/CGTGTGCGTTCCCAACT	Missense G	C	A	P

Figure 3